Uncertainty modelled using Probability, applications of Bayes formula for conditional prob. Introduction to Bayesian Networks

Henning Christiansen

Datalogi
Roskilde Universitetscenter

Program of today

- Uncertainty, AI, and probability theory
- Probability: Short intro + exercises
- Bayes' formula and applications: Short intro + exercises
- Bayesian networks: Short intro + exercises

Uncertainty?

- (Our knowledge about) reality is very seldom 100% certain
- Lack of knowledge, imprecise knowledge, making judgment from partial knowledge, thus conclusion cannot be exact but may express "degree of" (un)certainty of alternative conclusions
- The best known, and scientifically most well-founded background is Probability Theory
- Here, only time to very brief introduction and few applications in AI
- (there are other, more or less ad-hoc weighting mechanisms applied in expert systems, etc. ...)

Probability theory

- ***Random variables*** (here: discrete; can also be continuous)
- Can take one out of a set of values as result of an experiments or observation ("event")
- Variable V may take values $\{x_1, x_2, ..., x_n\}$
- Each value has a certain ***probability***, $P(V=x_i) \in [0,1]$
- By definition $P(V=x_1)+...+P(V=x_n) = 1$.
- Important: Probability function P(V=...) is a ***mathematical definition***, which has nothing to do with "average of ..."
- However: Probabilities should reflect reality, e.g., be defined from statistics.... which is a different matter!

Operations on events: $\cup$ and $\cap$

Corresponds to set operations on events

$P(V=x_1 \cup V=x_2) = P(V=x_1)+p(V=x_2)$ if $x_1 \neq x_2$

$P(V=x_1 \cap V=x_2) = 0$ if $x_1 \neq x_2$, ... not interesting

More interesting when applied for different rand. var's

$P(V=x \cap W=y)$

requires joint distribution given (as math. def.).

We could (but do not) write as P(VW=(x,y)).

Important: No a priori relationship

$P(V=x \cap W=y) = P(V=x)$???? $P(W=y)$

Conditional probabilities

Informally P(A|B) means probability of event A given that B has occurred (been observed)

Example: P(red-haired | girl) which abbreviates

P(V=red-haired | W=girl)

where possible values are W∈{boy,girl}, V ∈ {...}

Definition:

$P(A|B) = P(A \cap B) / P(B)$

Fits with intuition of Prob ≈ Relative Frequency:

$P(rh \mid g) \approx \big(\#(rh \cap g) / \#(b \cup g)\big) / \big(\#g / \#(b \cup g)\big)$

$= \#(rh \cap g) / \#g$

Dependent and independent events

Definition: Random variables V and W are *indep't* if

$P(V=x \cap W=y) = P(V=x) \times P(W=y)$ for all x, y

Proposition: Two ran. var's V and W are indep't iff

$P(V=x \mid W=y) = P(V=x)$ for all x, y

Example: What do the following mean intuitively?

P(red-haired ∩ girl) = P(red-haired) × P(girl)

P(red-haired | girl) = P(red-haired)

Definition: Two random variables are *dependent* if the are not independent ;-)

Now you do the work

Exercises in section 2.1 + 2.2 of the note

"Examples and exercises for conditional probabilities and Bayesian reasoning"

If you have no done it already, start reading text of section 2.1

NB: Notice also new concept of "exhaustive set of events" and Bayes' formula (3.11), plus sum versions 3.12–13.

Bayesian reasoning

Bayes' formula: Twisting conditional probabilities

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Splitting up p(B) in two cases, conditioned with A and ¬A:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|\neg A) \times P(\neg A)}$$

Example with A=woman, B=read-haired ....

An example ...

A red-haired person is seen running away from scene of crime...

Police has two suspects in custody, both red-haired, a man and a woman.

Who did it (probably):

$$P(\text{woman}|\text{red}) = \frac{P(\text{red}|\text{woman}) \times P(\text{woman})}{P(\text{red}|\text{woman}) \times P(\text{woman}) + P(\text{red}|\text{man}) \times P(\text{man})}$$

Well, if we know P(man), P(woman), how many men are typically red-haired and do. for women...

Some definitions and clarification: ...

Prior and posteriory probabilities

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|\neg A) \times P(\neg A)}$$

Probabilities P(A) and P(¬A) are called ***prior*** probabilities as they refer to probabilites that are given before any event has been observed

Probability P(A|B) and P(¬A|B) are called ***posteriory*** probabilities as they are calculated after some event has been observed

Back to the example ...

Consider again:

$$P(\text{woman}|\text{red}) = \frac{P(\text{red}|\text{woman}) \times P(\text{woman})}{P(\text{red}|\text{woman}) \times P(\text{woman}) + P(\text{red}|\text{man}) \times P(\text{man})}$$

We may have P(woman)=0.6 and P(man)=0.4.

But add now "80% of all criminal are men"...

changes P(woman)=0.2 and P(man)=0.8, so

with new prior probabilities, new posteriori are calculated...

Now you do the work

Exercises in section 3 of the note

"Examples and exercises for conditional probabilities and Bayesian reasoning"

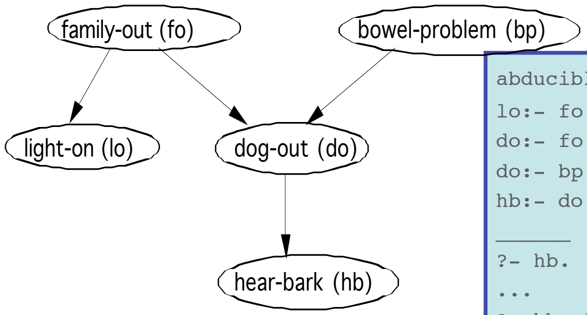
If you have not done it already, start reading text of section 3

1. Work with the red-haired woman/man example
2. More natural example about medical tests

Bayesian networks

- Conditional prop's ≈ logical rules
- P(*effect*|*cause*) ≈ *effect* :- *cause*.
- easier to measure than P(*cause*|*effect*)
- Bayesian network: Graph (DAG) of cause-effect relationships
  - ≈ a logic program
  - with limited structure and no arguments
  - but with probabilities
- Here: Discrete BNs
  - examples even binary = boolean
  - but any finite no. of possible outcomes of each random variables

Example (Charniak)



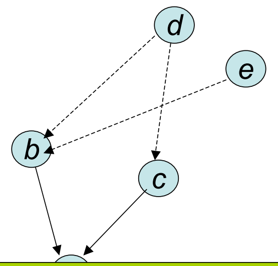

```
abducibles fo/0, bp/0.
lo:- fo.
do:- fo.
do:- bp.
hb:- do.
______
?- hb.
...
?- hb, lo.
```


(This and another figure copied from slides by J.-C. Latombe; found at http://www.cs.ualberta.ca/~lindek/366/)

Purpose of BN

- Statistically based, *abductive* reasoning, i.e., reasoning from "observed effect" to "(hidden) causes" with probabilities
- Based on conditional probabilities and Bayes' theorem ≈ a way of "reasoning backwards" in conditional probs.

Assumption of independence



a not necessarily indept. of *d* and *e*!!

... but cond. prob. are:

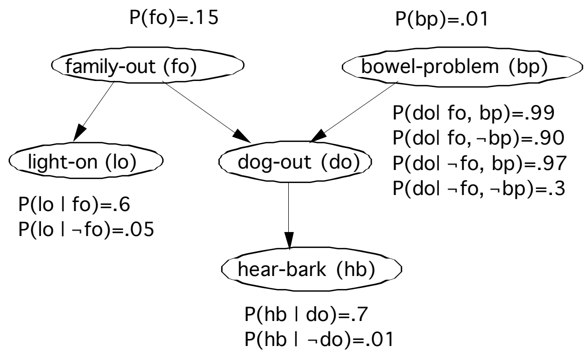
$$P(a|b,c) = P(a|b,c,d,e)$$

Intuitively:

a depends on actual values of *b* and *c*, but not on ***why*** *b* and *c*

You may try to read def. of "d-connected", but you're not expected to be able to reproduce it ;-)

Adding conditional probabilities

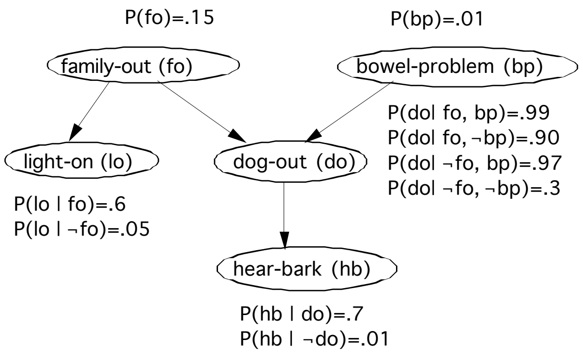


Notice:

- P(lo|fo) stands for P(lo=true|fo=true)
- P(lo|fo)=0.6 indicates implicitly

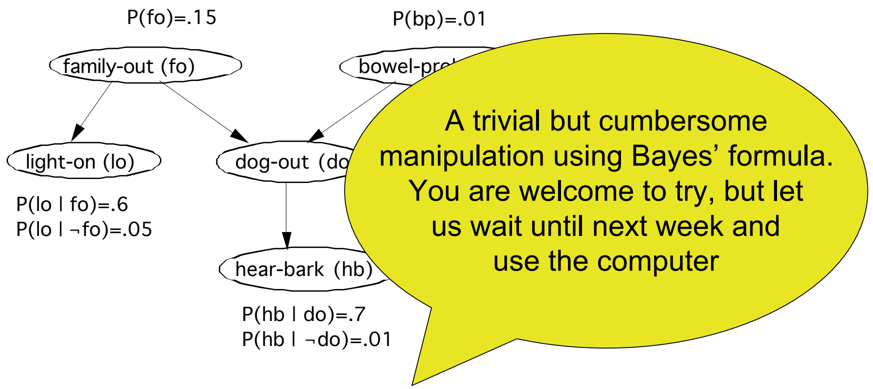
  P(not lo|fo)=P(lo=false|fo=true)=0.6

A little exercise



- Given fo=true and bp=false, calculate probability for P(hb=true)

Another little exercise



- Given P(hb=true), calculate probabilities for fo, bp

You exercise:

Exercise 4.1 in the note for today

- design a Bayesian network for the familiar power supply example

Exercise 4.2 (discussion; if time)

- on “intelligent” but annoying systems